

PATENT ABSTRACTS OF JAPAN

(11) Publication number : 08-328591

(43) Date of publication of application : 13.12.1996

(51)Int.CI. G10L 7/02
G10L 9/14
H03H 21/00
H03M 7/30

(21)Application number : 08-123685
(22)Date of filing : 17.05.1996

(71)Applicant : FR TELECOM

(72) Inventor : STEPHANE PROUST

(30)Priority

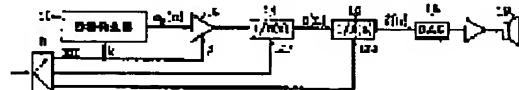
Priority number : 95 9505851 Priority date : 17.05.1995 Priority country : FR

(54) METHOD FOR APPLYING NOISE MASKING LEVEL TO COMPOSED ANALYTICAL VOICE CODER USING SHORT-PERIOD PERCEPTION WEIGHTING FILTER

(57)Abstract:

PROBLEM TO BE SOLVED: To apply a noise masking level to a composed analytical voice coder using a short-period perception weighting filter.

SOLUTION: In a composed analytical voice coder using a short-period perception weighting filter having a transmission function of $W(z) = A(z/\gamma_1)/A(z/\gamma_2)$, spectrum extension coefficients γ_1 and γ_2 are dynamically applied on the basis of a spectral parameter obtained in short-period linear predicting analysis. The spectral parameter useful for this application may include, particularly, a parameter showing all inclinations of the spectrum of a voice signal and a parameter showing the resonance characteristic of a short-period composing filter 16.



LEGAL STATUS

[Date of request for examination] 02.11.2000

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平8-328591

(43)公開日 平成8年(1996)12月13日

(51)Int.Cl. ⁶	識別記号	府内整理番号	F I	技術表示箇所
G 10 L 7/02			G 10 L 7/02	A
9/14			9/14	C
H 03 H 21/00	8842-5J		H 03 H 21/00	A
H 03 M 7/30	9382-5K		H 03 M 7/30	Z
審査請求 未請求 請求項の数7 OL (全10頁)				

(21)出願番号 特願平8-123685

(22)出願日 平成8年(1996)5月17日

(31)優先権主張番号 95 05851

(32)優先日 1995年5月17日

(33)優先権主張国 フランス (FR)

(71)出願人 591034154

フランス・テレコム

FRANCE TELECOM

フランス国、75015 パリ、プラス・ダル
レ、6

(72)発明者 ステファン・ブルースト

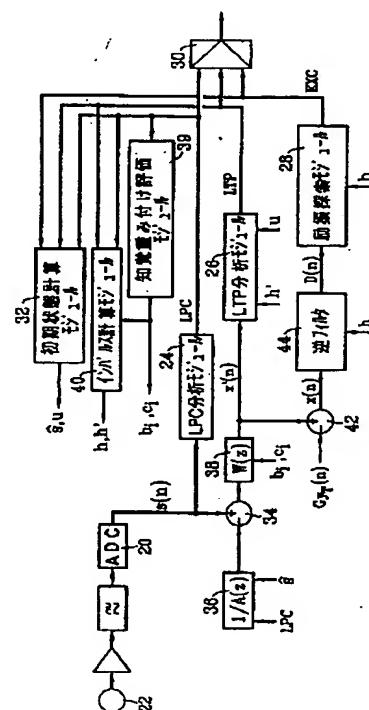
フランス・22300・ランニヨン・アレ
デ・ローズ・レジダンス・シャルコー (番
地なし)

(74)代理人 弁理士 志賀 正武 (外2名)

(54)【発明の名称】 短期知覚重み付けフィルタを使用する合成分析音声コーダに雑音マスキングレベルを適応する方
法

(57)【要約】

【課題】短期知覚重み付けフィルタを使用する合成分析音声コーダに雑音マスキングレベルを適応する。

【解決手段】伝達関数 $W(z) = A(z/\gamma_1)/A(z/\gamma_2)$ を有する短期知覚重み付けフィルタを使用する合成分析音声コーダでは、スペクトル拡張係数 γ_1 及び γ_2 は、短期線形予測分析中に得られたスペクトラパラメータに基づいて動的に適応させられる。この適応に役立つスペクトルパラメータは、特に、音声信号のスペクトルの全傾斜を表すパラメータ及び短期合成フィルタの共振特性を表すパラメータを含んでもよい。

【特許請求の範囲】

【請求項 1】 合成分析音声符号化方法において、短期合成フィルタ (16) を規定するパラメータ (LPC) を決定するために連続フレームとしてデジタル化される音声信号 ($s(n)$) の次数 p の線形予測分析ステップと、前記音声信号を表す合成信号を生成するための短期合成フィルタに適用される励振信号を規定する励振パラメータの決定ステップであって、前記励振パラメータの少なくともいくつかが、その伝達関数が式 $W(z) = A(z/\gamma_1)/A(z/\gamma_2)$ であるところの少なくとも一つの知覚重み付けフィルタによって前記音声信号と前記合成信号との差のフィルタリングの結果生じる誤差信号のエネルギーを最小にすることによって決定されることと、ここで、

$$A(z) = 1 - \sum_{i=1}^p a_i z^{-i}$$

係数 a_i は線形予測分析ステップで得られた線形予測係数であり、かつ γ_1 及び γ_2 は $0 \leq \gamma_2 \leq \gamma_1 \leq 1$ であるようなスペクトル拡張係数を示し、

前記短期合成フィルタを規定するパラメータ及び励振パラメータの量子化値の生成ステップとからなり、前記スペクトル拡張係数の少なくとも一つの値が前記線形予測分析ステップで得られたスペクトルパラメータに基づいて適応させられることを特徴とする合成分析音声符号化方法。

【請求項 2】 前記スペクトル拡張係数の少なくとも一つの値がそれに基づいて適応される前記スペクトルパラメータが、前記音声信号のスペクトルの全傾斜を表す少なくとも一つのパラメータ (r_1, r_2) と前記短期合成フィルタ (16) の共振特性を表す少なくとも一つのパラメータ (d_{min}) を含むことを特徴とする請求項 1 による方法。

【請求項 3】 前記スペクトルの全傾斜を表す前記パラメータが、前記線形予測分析中に決定される第 1 及び第 2 の反射係数 (r_1, r_2) を含むことを特徴とする請求項 2 による方法。

【請求項 4】 前記共振特性を表す前記パラメータが、2 つの連続ラインスペクトル周波数間の距離の最小 (d_{min}) であることを特徴とする請求項 2 又は 3 による方法。

【請求項 5】 いくつかのクラス (P_0, P_1) の中の音声信号のフレームの分類が前記スペクトルの全傾斜を表すパラメータ (r_1, r_2) に基づいて実行され、かつ各クラスに対して、2 つのスペクトル拡張係数が、前記短期合成フィルタ (16) の共振特性が増加するにつれてそれらの差 $\gamma_1 - \gamma_2$ が減少するように選定されることを特徴とする請求項 2 ないし 4 のいずれかによる方法。

【請求項 6】 第 1 の反射係数 $r_1 = R(1)/R(0)$ の値及び第 2 の反射係数 $r_2 = [R(2) - r_1 \cdot R(1)]/[(1 - r_1^2) \cdot R$

(0)] の値に基づいて選択された 2 つのクラスが提供され、 $R(j)$ が j 個のサンプルの遅延のための音声信号の自動相関関係を示し、かつ前記第 1 の反射係数 (r_1) が第 1 の正のしきい値 (T_1) よりも大きく、かつ前記第 2 の反射係数 (r_2) が第 1 の負のしきい値 ($-T_2$) よりも小さい第 1 のクラス (P_1) が各フレームから選択され、前記第 1 の反射係数 (r_1) が前記第 1 の正のしきい値よりも小さい第 2 の正のしきい値 (T_1') よりも小さく又は前記第 2 の反射係数 (r_2) が前記第 1 の負のしきい値 ($-T_2$) よりも絶対値で小さい第 2 の負のしきい値 ($-T_2'$) よりも大きい第 2 のクラス (P_0) が各フレームから選択されることを特徴とする請求項 3 ないし 5 のいずれかによる方法。

【請求項 7】 各クラス (P_0, P_1) において、前記スペクトル拡張係数の最大 γ_1 が固定され、かつ前記スペクトル拡張係数の最小 γ_2 が 2 つの連続ラインスペクトル周波数間の距離の最小 (d_{min}) の減少アフィン関数であることを特徴とする請求項 4 又は 5 による方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、合成による分析の技術を使用する音声の符号化に関するものである。

【0002】

【従来の技術】 合成音声符号化分析方法は、通常下記のステップからなる。短期合成フィルタを規定するパラメータを決定するために連続フレームとしてデジタル化される音声信号の次数 p の線形予測分析ステップ、音声信号を表す合成信号を生成するための短期合成フィルタに印加される励振信号を規定する励振パラメータの決定ステップであって、この励振パラメータの少なくともいくつかは、音声信号と少なくとも一つの知覚重み付けフィルタによる合成信号との差のフィルタリングの結果生じる誤差信号のエネルギーを最小にすることによって決定されていること、短期合成フィルタを規定するパラメータの量子化値と励振パラメータの量子化値の発生ステップ。

【0003】 線形予測によって得られる短期合成フィルタのパラメータは、声道及び入力信号のスペクトル特性の伝達関数を表す。いろいろなクラスの合成分析コーダを区別できる、短期合成フィルタに印加される励振信号をモデル化するいろいろな方法がある。存在する大方のコーダでは、励振信号は、声帯の振動による母音のような有音の長期周期性を利用して適応コードブックによるか又は長期合成フィルタによって合成された長期成分を含んでいる。CELPコーダ ("符号励振線形予測 (Code Excited Linear Prediction)", エム・アール・シュローダー (M. R. Schroder) 及びビー・エス・アタル (B. S. Atal) 著の "符号励振線形予測 (Code Excited Linear Prediction) (CELP): 非常に低いビット速度での高品質音声 (High Quality Speech at Very Low Bit Rates)", Proc. ICA

SSP' 85, Tampa, March 1985, ページ937~940を参照)では、誤差励振は、確率コードブックから抽出される波形によってモデル化され、利得と乗算される。CELPコーダは、通常の電話帯域において、音声の品質をそこなわないで、64キロビット/秒(従来のPCMコーダ)から16キロビット/秒(LD-CELPコーダ)まで及び大部分の最新のコーダに対しては8キロビット/秒以下にさえ必要とされるディジタルビット速度を減少することを可能にしている。これらのコーダは、今日では一般に電話伝送で使用されるが、記憶、広帯域電話又は衛星伝送のような多数の他のアプリケーションを提供する。本発明が適用される合成分析コーダの他の例としては、特に、MP-LPCコーダ(マルチパルス線形予測コーディング(Multi-Pulse Linear Predictive Coding), ピー・エス・アタル(B. S. Atal)及びジー・アール・レミデ(J. R. Remde)著の“低ビット速度での自然に聞こえる音声を発生するためのLPC励振の新しいモデル”(A New Model of LPC Excitation for Producing Natural-Sounding Speech at Low Bit Rates), Proc. ICASSP' 82, Paris, May 1982, Vol. 1, ページ614~617を参照)があげられ、この場合、誤差励振はそれに割り当てられたそれぞれの利得を有する可変位置パルスによってモデル化されており、他の例としてはVSELPコーダ(ベクトル和励振線形予測(Vector-Sum Excited Linear Prediction), アイ・エー・ガーソン(I. A. Gerson)及びエム・エー・ジャーシウク(M. A. Jasiuk)著の“8キロビット/秒でのベクトル和励振線形予測(VSELP)音声符号化(Vector-Sum Excited Linear Prediction(VSELP) Speech Coding at 8kbits/s)”, Proc. ICASSP' 90 Albuquerque, April 1990, Vol. 1, ページ461~464を参照)があり、この場合、励振がそれぞれのコードブックから抽出されたパルスベクトルの線形結合によってモデル化される。

【0004】コーダは、合成信号と最初の音声信号との間の知覚的に重み付けられた誤差を最小にする“閉ループ”処理において誤差励振を評価する。知覚重み付けが、平均二乗誤差の直接最小化に関しては、合成音声の主観的知覚を実質的に改善することが公知である。短期知覚重み付けは、最小誤差判定基準内で信号レベルが比較的高い重要な音声スペクトルの領域を減ずることである。すなわち、聞き手によって知覚される雑音は、フラットであったスペクトルがフォルマント間の領域内よりもフォルマント領域内により多くの雑音を受け取るような形状にされるならば、減少される。これを達成するために、短期知覚重み付けフィルタは、下記の式の伝達関数をしばしば有する。

$$W(z) = A(z) / A(z/\gamma)$$

ここで、

$$A(z) = 1 - \sum_{i=1}^p a_i z^{-i}$$

で、係数 a_i は線形予測解析ステップで得られる線形予測係数であり、かつ γ は0と1との間にあるスペクトル拡

張係数を示す。この式の重み付けは、ピー・エス・アタル(B. S. Atal)及びエム・アール・シューローダー(M. R. Schroeder)著の“音声信号の予測コーディング及び主観的誤差判定基準(Predictive Coding of Speech Signals and Subjective Error Criteria)”, IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-27, No. 3, June 1979, ページ247~254によって提案されている。 $\gamma=1$ に関しては、いかなるマスキングもない。すなわち、二乗誤差の最小化は、合成信号に基づいて実行される。 $\gamma=0$ であるならば、マスキングは完全である。すなわち、最少化は誤差に基づいて実行され、符号化雑音は音声信号と同一のスペクトル包絡を有する。

【0005】下記の式の伝達関数 $W(z)$ を知覚重み付けのために選択することで一般化することができる。

$$W(z) = A(z/\gamma_1) / A(z/\gamma_2)$$

γ_1 及び γ_2 は、 $0 \leq \gamma_2 \leq \gamma_1 \leq 1$ であるようなスペクトル拡張係数を示している。ジェー・エッチ・チーン(J. H. Chen)及びエー・ガーショウ(A. Gersho)著の“適応後フィルタリングを有する4800Bpsでの実時間ベクトルAPC音声コーディング(Real-Time Vector APC Speech Coding at 4800 Bps with Adaptive Postfiltering)”, Proc. ICASSP' 87, April 1987, ページ2185~2188を参照のこと。 $\gamma_1 = \gamma_2$ であるとき、マスキングがなくて、 $\gamma_1 = 1$ 及び $\gamma_2 = 0$ であるとき、マスキングが完全であることに注目されたい。スペクトル拡張係数 γ_1 及び γ_2 は、所望の雑音マスキングのレベルを決定する。マスキングがあまりにも弱いと、一定の粗い量子化雑音が知覚される。マスキングがあまりにも強いとフォルマントの形状に影響を及ぼし、したがってひずみが非常に聞こえるようになる。

【0006】大部分の強力な最新のコーダでは、LTP遅延及びおそらく位相(わずかな遅延)を含む長期予測値のパラメータ又は係数セット(マルチタップLTPフィルタ)もまた、知覚重み付けフィルタを含む閉ループ手順によって各フレーム又はサブフレームの間に決定される。あるコーダでは、音声信号の短期モデル化を利用し、雑音のフォルマント分布を提供する知覚重み付けフィルタは、高調波に対応するピークにおける雑音のエネルギーを増加し、これらのピーク間で雑音のエネルギーを減少させる調波重み付けフィルタ及び/又は特に広帯域アプリケーションにおいて高周波でマスクされない雑音の出現を防止するように意図された傾斜補正フィルタで補われる。

【0007】

【課題を解決するための手段】本発明は主に、短期知覚重み付けフィルタ $W(z)$ に関する。短期知覚フィルタのスペクトル拡張パラメータ γ 、すなわち γ_1 又は γ_2 の選択は、通常、主観的検査の助けを借りて最適化される。この選択はその後固定される。しかしながら、出願人は、入力信号のスペクトル特性によって、スペクトル拡張パラメータの最適値がかなり大きな変動を受け

得ることを観察している。したがって、なされた選択は多少は満足する妥協という性格のものである。本発明の目的は、知覚重み付けフィルタのより良い特性によって符号化信号の主観的品質を向上させることにある。他の目的は、いろいろな種類の入力信号に対してコーダの性能をより均一にすることにある。他の目的は、この改良に対して更に複雑なことを必要としないことにある。

【0008】本発明は、このように、知覚重み付けフィルタが前述のように一般式 $W(z) = A(z/\gamma_1)/A(z/\gamma_2)$ の伝送関数を有し、かつスペクトル拡張係数 γ_1 、 γ_2 の少なくとも一つの値が線形予測分析ステップで得られたスペクトルパラメータに基づいて適応される、最初に示したタイプの合成解析音声符号化方法に関するものである。知覚重み付けフィルタの係数 γ_1 及び γ_2 を適応できるようにすることによって、音声ピックアップの特性、音声のいろいろな特性又は顕著な背景雑音（例えば、移動無線電話における自動車雑音）に依存するかなり大きい変動を有し得る入力信号のいろいろなスペクトル特性に対して符号化雑音マスキングレベルを最適化することが可能になる。知覚される主観的品質は向上し、コーダの性能はいろいろな種類の入力に対してより均一になる。

【0009】それに基づきスペクトル拡張係数の少なくとも一つの値が適応されるスペクトルパラメータが、音声信号のスペクトルの全ての傾斜を表す少なくとも一つのパラメータを含むことが好ましい。音声スペクトルは、低周波（大人の男性の太い音声の60Hzから子供の音声の500Hzに及ぶおおよそ基本の周波数）で平均してより多いエネルギーを有し、それゆえに一般に下向きの傾斜を有する。しかしながら、大人の男性の太い音声は、より減衰された高周波を有し、したがってより大きい傾斜のスペクトルを有する。音声ピックアップ系によって適用される前フィルタリングはこの傾斜に大きな影響を有する。従来の電話送受話器は、この傾斜効果をかなり減衰する、IRSと呼ばれるハイパス前フィルタリングを実行する。しかしながら、それよりも最新のある装置でなされる“線形”入力は重要な低周波の全てを保持する。弱いマスキング（ γ_1 と γ_2 との間の小さい間隙）は、信号の傾斜と比較して知覚フィルタの傾斜を減衰し過ぎる。信号がこれらの周波数でほとんどエネルギーを有しないならば、高周波での雑音レベルは大きなままであり、信号そのものよりも大きくなる。耳は、高周波のマスクされていない雑音を知覚し、それはしばしば高調波特性を持っているのでなおさら耳障りである。このエネルギー差を適当にモデル化するためにはフィルタの傾斜の簡単な補正では適当でない。この問題は音声スペクトルの全傾斜を考慮に入れたスペクトル拡張係数の適応によって、より良く処理することができる。それに基づきスペクトル拡張係数の少なくとも一つの値が適応されるスペクトルパラメータが、短期合成フィルタ（LP

C）の共振特性を表す少なくとも一つのパラメータをさらに含むことが好ましい。音声信号は、電話帯域における4個又は5個までのフォルマントを有する。スペクトルの輪郭線を特徴付けるこれらの“こぶ”は、一般にかなり丸くされる。しかしながら、LPC分析は、不安定性に近いフィルタになることもある。したがって、LPCフィルタに対応するスペクトルは、小さい帯域幅にわたって大きなエネルギーを有する、比較的はっきりとしたピークを含む。マスキングが大きくなればなるほど、雑音のスペクトルはLPCスペクトルに接近する。しかしながら、雑音分布におけるエネルギーピークの存在は非常にやっかいである。エネルギーピークの存在によって、かなりのエネルギー範囲内でフォルマントレベルで歪みが生じ、それはかなり耳障りである。したがって、本発明は、LPCフィルタの共振特性が増加するにつれてマスキングのレベルを減少することを可能にするものである。

【0010】短期合成フィルタがラインスペクトルパラメータ又はラインスペクトル周波数（LSP又はLSF）によって表されるとき、 γ_1 及び/又は γ_2 のどの値に基づいて適応される、短期合成フィルタの共振特性を表すパラメータは、2つの連続ラインスペクトル周波数間の差の最小であり得る。

【0011】

【発明の実施の形態】本発明の他の特徴及び利点は、添付図面を参照しながら好ましいが限定していない実例の実施例に関する下記の説明で明かになる。本発明は、CELP形の音声コーダへのその適用において下記に示されている。しかしながら、本発明はまた、他の種類の合成分析コーダ（MP-LPC、VSELP...）に適用できることも理解される。CELPコーダ及びCELPデコーダで実施される音声合成処理が、図1に示されている。励振発生器10は、インデックス k に応じて所定のコードブックに属する励振コード c_k を供給する。增幅器12は、この励振コードと励振利得 β を乗算し、この結果得られる信号は長期合成フィルタ14に委ねられる。フィルタ14からの出力信号 u は順に短期合成フィルタ16に委ねられ、それからの出力 s は、ここでは合成音声信号とみなされるものを構成する。もちろん、他のフィルタ、例えば、後置フィルタも音声符号化の分野で周知であるように、デコーダレベルで具備することができる。

【0012】前述の信号は、例えば8kHzに等しいサンプリング速度 F_e で例えば16ビットによって表されるデジタル信号である。合成フィルタ14、16は、一般に純粋な再帰型フィルタである。長期合成フィルタ14は、一般に $B(z) = 1 - Gz^{-T}$ を有する式 $1/B(z)$ の伝達関数を有する。遅延 T 及び利得 G は、コーダによって適応できるように決定されている長期予測（LTP）パラメータを構成する。短期合成フィルタ16のLPCパラメータは、音声信号の線形予測によってコーダで決定される。従って、フィルタ16の伝達関数は、下記の式を有する式

$1/A(z)$ である。

$$A(z) = 1 - \sum_{i=1}^p a_i z^{-i}$$

次数 p (一般に $p=10$) の線形予測の場合、 a_i は i 番目の線形予測係数を表す。ここで、“励振信号”は、短期合成フィルタ14に印加される信号 $u(n)$ を示す。この励振信号は、LTP成分 $G \cdot u(n-T)$ 及び誤差成分、すなわち刷新シーケンス $\beta c_k(n)$ を含んでいる。合成分析コーダでは、誤差成分を特徴付けるパラメータ及び、任意にはLTP成分が、知覚重み付けフィルタを使用して閉ループで評価される。図2は、CELPコーダの配置図を示している。音声信号 $s(n)$ はデジタル信号で、例えば、マイクロホン22の増幅され、かつフィルタリングされた出力信号を処理するアナログ/デジタル変換器20によって供給される。信号 $s(n)$ は、それ自身を Δ 個のサンプルのサブフレーム、すなわち励振フレームに分割される Δ 個のサンプルの連続フレームとしてデジタル化される (例えば、 $\Delta=240$ 、 $L=40$)。

【0013】LPC、LTP及びEXCパラメータ (インデックス k 及び励振利得 β) は、3つのそれぞれの分析モジュール24、26、28によってコーダレベルで得られる。次に、これらのパラメータは、有効デジタル伝送のために公知の方法で量子化され、コーダからの出力信号を形成するマルチプレクサ30に委ねられる。これらのパラメータはコーダの特定のフィルタの初期状態を計算するためにモジュール32にも供給される。このモジュール32は本来、図1で表されるような復号化チェーンを含む。デコーダと同様に、モジュール32は、量子化LPC、LTP及びEXCパラメータに基づいて作動する。LPCパラメータの補間が一般に行われるようデコーダで実行されるならば、同一の補間がモジュール32によって実行される。モジュ

$$a_j^{(i)} = r_j \\ E(i) = (1 - r_j^2) \cdot E(i-1)$$

1から $i-1$ までとる j に関しては、下記のことをする。

$$a_j^{(i)} = a_j^{(i-1)} - r_j \cdot a_{j-1}^{(i-1)}$$

係数 $a_j^{(i)}$ は、最新の反復で得られる $a_j^{(p)}$ に等しく選ばれる。物理量 $E(p)$ は残留予測誤差のエネルギーである。 -1 と 1 の間にある係数 r_j は反射係数と呼ばれる。それらは、しばしば $10g$ 面積比 $LAR_j = LAR(r_j)$ によって表され、関数 LAR は、 $LAR(r) = 10g_{10} [(1-r)/(1+r)]$ によって規定される。

【0015】LPCパラメータの量子化は、直接に係数 a_j にわたって、反射係数 r_j にわたって又は $10g$ 面積比 LAR_j にわたって実行することができる。他の可能性は、ラインスペクトルパラメータを量子化することであ

ール32は、考慮中のサブフレームよりも先に合成及び励振パラメータに基づいて決定されるデコーダの合成フィルタ14、16の初期の状態の情報をコーダレベルで供給する。符号化処理の第1のステップでは、短期分析モジュール24は、音声信号 $s(n)$ 短期相関を分析することによってLPCパラメータ (短期合成フィルタの係数 a_j) を決定する。この決定は、音声信号のスペクトル内容の変化に適応させるように、 Δ サンプルのフレーム毎に例えれば一度実行される。LPC分析法は当該技術分野で周知である。例えば、1978年、ブレンティスホール社発行のエル・アール・ラビナー (L.R. Rabiner) 及びアール・ダブリュー・シャファー (R.W. Shafer) 著による文献 “音声信号のデジタル処理 (Digital Processing of Speech Signals)” を参照してもよい。この文献は、特に下記のステップを含むダービンのアルゴリズムを記載している。

【0014】フレームの長さが小さいならば (例えば、 $20 \sim 30$ ms)、現在のフレーム及びおそらくそれよりも前のサンプルを含む分析ウィンドウにわたって音声信号 $s(n)$ の p 個の自動相関関係 $R(i)$ ($0 \leq i < p$) の評価ステップ: $M \geq \Delta$ 及び $s^*(n) = s(n) \cdot f(n)$ を有する。

M-1

$$R(i) = \sum_{n=1}^{M-1} s^*(n) \cdot s^*(n-i)$$

$f(n)$ は、長さ M のウィンドウ関数、例えば矩形関数又はハミング関数を示す。

係数 a_j の再帰評価ステップ:

$$E(0) = R(0)$$

1から p までとる i に関しては、下記のことをする。

$$r_i = [R(i) - \sum_{j=1}^{i-1} a_j^{(i-1)} \cdot R(i-j)] / E(i-1)$$

る (LSPは “ラインスペクトル対” を表し、LSFは “ラインスペクトル周波数” を表す)。0と π との間で正規化された p 個のラインスペクトル周波数 ω_j ($1 \leq i \leq p$) は、複素数 $1, \exp(j\omega_2), \exp(j\omega_4), \dots, \exp(j\omega_p)$ が、多項式 $P(z) = A(z) - z^{-(p+1)} A(z^{-1})$ の平方根であり、複素数 $\exp(j\omega_1), \exp(j\omega_3), \dots, \exp(j\omega_{p-1})$ 及び -1 が、多項式 $Q(z) = A(z) + z^{-(p+1)} A(z^{-1})$ の平方根であるようなものである。量子化は、正規化周波数 ω_j 又はその余弦によって実行することができる。モジュール24は、本発明を実施する際に有用である物理量 r_j 、 LAR_j 及び ω_j を規定するために上述されたダービンの古典アルゴリズムによりLPC分析を実行できる。より最近に開発された同一結果を与える他のアルゴリズム、特にレビンソンのスプリットアルゴリズム (エス・サオウディ (S. Saoudi)、

ジェー・エム・ボウチャー(J. M. Boucher)及びエー・レーガイダー(A. Le Guyader)著の“音声符号化のためのLSPパラメータを計算するための新しい有効なアルゴリズム”(A new Efficient Algorithm to Compute the LSP Parameters for Speech Coding), Signal Processing, Vol. 28, 1992, ページ201~212を参照)又はチェビシェフの多項式(ピー・キャバル(P. Kabal)及びアル・ピー・ラマシヤンドラ(R. P. Ramachandran)著の“チェビシェフ多項式を使用するラインスペクトル周波数の計算”, IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. 1, ASSP-34, No. 6, ページ1419~1426, December 1986を参照)を都合よく使用することができる。

【0016】符号化の次のステップは長期予測LTPパラメータを決定することである。例えば、L個のサンプルのサブフレーム毎に一度決定される。減算器34は、マル入力信号に対する短期合成フィルタ16の応答を音声信号 $s(n)$ から減算する。この応答は伝達関数 $1/A(z)$ を有するフィルタ36によって決定され、その係数はモジ

$0 < i \leq p$ に対して $c_0 = 1$ 及び $c_i = -a_i \gamma_1^i$ である場合は、下記の伝達関数を有する次数 p の全ゼロ点の連続する級数とみなすことができる。

$$A(z/\gamma_1) = \sum_{i=0}^p c_i z^{-i}$$

このように、モジュール39は、各フレームに対する係数

$$G = \frac{\sum_{n=0}^{L-1} x'(n) \cdot y_T(n)}{\sum_{n=0}^{L-1} [y_T(n)]^2}$$

ここで、 $x'(n)$ は、関連サブフレームの間のフィルタ38からの出力信号を示し、 $y_T(n)$ は、畳み込み積 $u(n-T) * h'(n)$ を示す。上記の式では、 $h'(0)$ 、 $h'(1)$ 、 \dots 、 $h'(L-1)$ は、伝達関数 $W(z)/A(z)$ を有する重み付け合成フィルタのインパルス応答を示している。このインパルス応答 h' は、量子化及び補間後に必要とされるならば、モジュール39によって供給される係数 b_i 及び c_i とサブフレームのために決定されるLPCパラメータに基づいて、インパルスを計算するモジュール40によって得られる。サンプル $u(n-T)$ は、モジュール32によって供給されるような長期合成フィルタ14の初期状態である。サブフレームの長さよりも小さい遅延 T に関しては、欠けているサンプル $u(n-T)$ は、初期の

$$G = \frac{\sum_{n=0}^{L-1} x'(n) \cdot y_T(n)}{\sum_{n=0}^{L-1} [y_T(n)]^2}$$

【0017】サブフレームに関するCELP励振を探索するために、最適遅延 T に関してモジュール26で計算された信号 $Gy_T(n)$ は、まず減算器42によって信号 $x'(n)$ か

ユール24によって決定されたLPCパラメータによって与えられ、かつその初期状態 s が合成信号の最後の p 個のサンプルに対応するようにモジュール32によって供給される。減算器34からの出力信号は、その役割が誤差が最も知覚できるスペクトルの一部、すなわちフォルマント間領域を強調することである知覚重み付けフィルタに委ねられる。知覚重み付けフィルタの伝達関数 $W(z)$ は、一般式 $W(z) = A(z/\gamma_1)/A(z/\gamma_2)$ であり、 γ_1 及び γ_2 は、 $0 \leq \gamma_2 \leq \gamma_1 \leq 1$ であるような2つのスペクトル拡張係数である。本発明は、LPC分析モジュール24によって決定されたスペクトルパラメータに基づいて γ_1 及び γ_2 の値を動的に適応させることを提案する。この適応は、さらに記載してある処理により、知覚重み付けを評価するモジュール39によって実行される。知覚重み付けフィルタは、 $0 < i \leq p$ に対して $b_0 = 1$ 及び $b_i = -a_i \gamma_2^i$ である場合は、下記の伝達関数を有する次数 p の全極点の連続する級数とみなすことができ、

$$1/A(z/\gamma_2) = 1 / \left[\sum_{i=0}^p b_i z^{-i} \right]$$

b_i 及び c_i を計算し、これらをフィルタ38に供給する。モジュール26によって実行される閉ループLPT分析は、下記の正規化された相関関係を最大にする遅延 T を従来のように各サブフレームに対して選択するものである。

サンプルに基づいて補間によって得られるか又は音声信号から得られる。整数又は分数である遅延 T は、例えば20のサンプルから143までのサンプルに及ぶ指定ウィンドウから選択される。閉ループ探索範囲を減少する、したがって計算される畳み込み $y_T(n)$ の数を減少するために、フレーム毎に例えば1回閉ループ遅延 T' を決定し、次に、減少された間隔約 T' で各サブフレームに対して閉ループ遅延を選択することがまず可能であろう。閉ループ探索はもっと単純に、伝達関数 $A(z)$ を有する逆フィルタによって多分フィルタリングされる音声信号 $s(n)$ の自動相関関係を最大にする遅延 T' を決定することである。一旦遅延 T が決定されると、長期予測利得 G は下記によって得られる。

$$G = \frac{\sum_{n=0}^{L-1} x(n) \cdot y_T(n)}{\sum_{n=0}^{L-1} [y_T(n)]^2}$$

ら減算される。得られた信号 $x(n)$ は、逆フィルタ44に委ねられ、逆フィルタ44は下記の式で表される信号を提供する。

$$D(n) = \sum_{i=n}^{L-1} x(i) \cdot h(i-n)$$

ここで、 $h(0)$ 、 $h(1)$ 、 \dots 、 $h(L-1)$ は、合成フィルタ及び知覚重み付けフィルタから成る複合フィルタのインパルス応答を示し、この応答はモジュール40によって計算される。すなわち、複合フィルタは、伝達関数 $W(z)/A(z) \cdot B(z)$ を有する。したがって、マトリックス表示では、下記のような式を得る。

$$x = (x(0), x(1), \dots, x(L-1)) \text{の場合,}$$

$$D = (D(0), D(1), \dots, D(L-1)) = x \cdot H$$

及び

【数1】

$$H = \begin{bmatrix} h(0) & 0 & \cdot & \cdot & 0 \\ h(1) & h(0) & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ h(L-2) & & h(0) & 0 \\ h(L-1) & h(L-2) & \cdot & h(1) & h(0) \end{bmatrix}$$

【0018】ベクトル D は、励振探索モジュール28のための目標ベクトルを構成する。このモジュール28は、下記のような正規化相関関係 P_k^2/α_k^2 を最大にするコードブックからのコード語を決定する。

$$P_k = D \cdot c_k^T$$

$$\alpha_k^2 = c_k \cdot H^T \cdot H \cdot c_k^T = c_k \cdot U \cdot c_k^T$$

最適インデックス k が決定されると、励振利得 β は $\beta = P_k/\alpha_k^2$ に等しくなるように取られる。図1を参照すると、CELPデコーダは、コーダによる2進ストリーム出力を受け取る多重分離装置8を備えている。EXC励振パラメータの量子化値とLTP合成パラメータ及びLPC合成パラメータの量子化値は合成信号 s を再構成するために発生器10、增幅器12及びフィルタ14、16に供給され、この合成信号は例えば、增幅される前に変換器18によってアナログ変換され、次に元の音声を復元するためにスピーカ19に印加されることができる。それに基づいて係数 γ_1 及び γ_2 が適応されるスペクトルパラメータは、一方では音声スペクトルの全傾斜を表す最初の2つの反射係数 $r_1 = R(1)/R(0)$ 及び $r_2 = [R(2) - r_1 R(1)] / [(1 - r_1^2) R(0)]$ と他方ではその分布が短期合成の共振特性を表すラインスペクトル周波数とを含む。短期合成フィルタの共振特性は、2つのラインスペクトル周波数間の最小距離 d_{min} が減少するにつれ増加する。周波数 ω_i は、昇順($0 < \omega_1 < \omega_2 < \dots < \omega_p < \pi$)で得られるので、下記の式を得る。

$$d_{min} = \min_{1 \leq i < p} (\omega_{i+1} - \omega_i)$$

【0019】前述のダービンのアルゴリズムの最初の反

復で中止することによって、音声スペクトルのおおまかな近似は、伝達関数 $1/(1 - r_1 \cdot z^{-1})$ によって生成される。したがって、合成フィルタの全傾斜（通常は負）は、第1の反射係数 r_1 が1に近づくにつれて絶対値で増加する傾向がある。分析が反復を付加することによって次数2まで続けられるならば、あまりおおまかでないモデル化が、伝達関数 $1/[1 - (r_1 - r_1 r_2) \cdot z^{-1} - r_2 \cdot z^{-2}]$ を有する次数2のフィルタで達成される。次数2のこのフィルタの低周波共振特性は、その極点がユニット円に近づくにつれて、すなわち r_1 が1に、 r_2 が-1に近づくにつれて増加する。したがって、音声スペクトルは、 r_1 が1に近づき、 r_2 が-1に近づくにつれて、低周波で比較的大きなエネルギー（言い換えれば比較的大きな負の全傾斜）を有すると結論付けることができる。音声スペクトルにおけるフォルマントピークはいくつかのラインスペクトル周波数（2又は3）を一緒に束にするのに対して、スペクトルの平たい部分はこれらの周波数の均一な部分に対応することは公知である。したがってLPCフィルタの共振特性は、距離 d_{min} が減少するにつれて増加する。一般に、合成フィルタのローパス特性が増加する（ r_1 は1に近づき、 r_2 は-1に近づく）につれて、及び／又は合成フィルタの共振特性が減少する（ d_{min} は増加する）につれて、より大きなマスキングが選定される（ γ_1 と γ_2 との間のより大きな間隙）。

【0020】図3は、知覚重み付けを評価するためにモジュール39が各フレームで実行する動作の典型的なフローチャートを示している。各フレームで、モジュール39は、モジュール24からLPCパラメータ a_i 、 r_i （又は LA_R_i ）及び ω_i （ $1 \leq i \leq p$ ）を受け取る。ステップ50では、モジュール39は、 $1 \leq i < p$ に対して $\omega_{i+1} - \omega_i$ を最小化することによって2つの連続するラインスペクトル周波数間の最小距離 d_{min} を評価する。フレームにわたるスペクトルの全傾斜を表すパラメータ（ r_1 及び r_2 ）に基づいて、モジュール39は、N個のクラス P_0 、 P_1 、 \dots 、 P_{N-1} の間のフレームの分類を実行する。図3の例では、 $N=2$ である。クラス P_1 は、音声信号 $s(n)$ が低周波で比較的効果的である（1に比較的近い r_1 及び-1に比較的近い r_2 ）場合に対応する。したがって、一般にクラス P_1 ではクラス P_0 で取り入れられるよりも大きなマスキングが取り入れられる。クラス間を極端に頻繁に遷移することを避けるために、いくらかのヒステリシスが r_1 及び r_2 の値に基づいて導入される。たとえばクラス P_1 に対して各フレームから r_1 が正のしきい値 T_1 よりも大きく、かつ r_2 が負のしきい値- T_2 よりも小さく、選択し、クラス P_0 に対して各フレームから r_1 が他の正のしきい値 T_1' よりも小さく（ $T_1' < T_1$ の場合）、かつ r_2 が他の負のしきい値- T_2' よりも小さく（ $T_2' < T_2$ の場合）、選択すると仮定する。反射係数約±1の感度を与えられると、このヒステリシスは、しきい値 T_1 、 T_1' 、 $-T_2$ 、 $-T_2'$ がそれぞれしきい値 $-S_1$ 、 $-S_1'$ 、 $-T_2$ 、 $-T_2'$ にそれぞれ対応する。

S_1' 、 S_2 、 S_2' に対応する \log 面積比LAR (図4を参照) の領域で容易に視覚化できる。初期設定の際に、デフォルトクラスは、例えば、マスキングが最も少ないクラス(P_0)である。ステップ52では、モジュール39は、前のフレームがクラス P_0 の下又はクラス P_1 の下にくるかどうかを調べる。前のフレームがクラス P_0 であるならば、モジュール39は、54で、条件 ($LAR_1 < -S_1$ 及び $LAR_2 > S_2$) をテストするか又はモジュール24が \log 面積比 LAR_1 、 LAR_2 の代わりに反射係数 r_1 、 r_2 を供給するならば、同等な条件 ($r_1 > T_1$ 及び $r_2 < -T_2$) をテストする。 $LAR_1 < -S_1$ 及び $LAR_2 > S_2$ ならば、クラス P_1 (ステップ56) に遷移する。テスト54が、 $LAR_1 \geq -S_1$ 又は $LAR_2 \leq S_2$ であることを示すならば、現在のフレームがクラス P_0 にとどまる (ステップ56)。

【0021】ステップ52が前のフレームがクラス P_1 であることを示すならば、モジュール39は、60で、条件 ($LAR_1 > -S_1'$ 又は $LAR_2 < S_2'$) をテストするか又はモジュール24が \log 面積比 LAR_1 、 LAR_2 の代わりに反射係数 r_1 、 r_2 を供給するならば、同等な条件 ($r_1 < T_1'$ 又は $r_2 > -T_2'$) をテストする。 $LAR_1 > -S_1'$ 又は $LAR_2 < S_2'$ ならば、クラス P_0 (ステップ58) に遷移する。テスト60が、 $LAR_1 \leq -S_1'$ 及び $LAR_2 \geq S_2'$ であることを示すならば、現在のフレームがクラス P_1 にとどまる (ステップ56)。図3で示される例では、2つのスペクトル拡張係数の大きい方の係数 γ_1 は、 $\Gamma_0 \leq \Gamma_1$ の場合、各クラス P_0 、 P_1 で定数値 Γ_0 、 Γ_1 を有し、他のスペクトル拡張係数 γ_2 は、ラインスペクトル周波数間の最小距離 d_{\min} の減少アフィン関数である。すなわち、 $\lambda_0 \geq \lambda_1 \geq 0$ 及び $\mu_1 \geq \mu_0 \geq 0$ の場合、クラス P_0 では $\gamma_2 = -\lambda_0 \cdot d_{\min} + \mu_0$ で、クラス P_1 では $\gamma_2 = -\lambda_1 \cdot d_{\min} + \mu_1$ である。 γ_2 の値はまた極端に急な変動を避けるために結合することもできる。すなわち、クラス P_0 では、 $\Delta_{\min,0} \leq \gamma_2 \leq \Delta_{\max,0}$ 、クラス P_1 では、 $\Delta_{\min,1} \leq \gamma_2 \leq \Delta_{\max,1}$ である。現在のフレームの間に選ばれたクラスに応じて、モジュール39は、ステップ56又は58で γ_1 及び γ_2 の値を割り当て、次に、ステップ62で知覚重み付け因数の係数 b_i 及び c_i を計算する。

【0022】前述のように、モジュール24がLPCパラメータを計算する Λ 個のサンプルのフレームは、励振信号を決定するために L 個のサンプルのサブフレームに細分割される。一般に、LPCパラメータの補間はサブフレームレベルで実行される。この場合、補間されたLPCパラメータを使って、各サブフレーム又は励振フレームに対して図3の処理を実施することが望ましい。出願人は、8キロビット/秒で作動する代数コードブックCELPコー

ダの場合に、そのためのLPCパラメータを各10msフレーム ($\Lambda=80$) で計算し、係数 γ_1 及び γ_2 を適応させる処理をテストした。フレームは、励振信号を探索するために2つの5msサブフレーム ($L=40$) にそれぞれ分割される。フレームのために得られたLPCフィルタは第2のこれらのサブフレームに対して適用される。第1のサブフレームに関しては、補間がこのフィルタと前のフレームの間に得られたフィルタとの間のLSE領域で実行される。マスキングレベルを適応させる手順は、LSF ω_i の補間及び第1のサブフレームに対する反射係数 γ_1 、 γ_2 の補間によってサブフレームの速度で適用される。図3で示される手順は、以下の数値とともに使用される。すなわち、 $S_1=1.74$ ； $S'_1=1.52$ ； $S_2=0.65$ ； $S'_2=0.43$ ； $\Gamma_0=0.94$ ； $\lambda_0=0$ ； $\mu_0=0.6$ ； $\Gamma_1=0.98$ ； $\lambda_1=6$ ； $\mu_1=1$ ； $\Delta_{\min,1}=0.4$ ； $\Delta_{\max,1}=0.7$ 、で周波数 ω_i は、0と π の間で正規化される。

【0023】余分な複雑なことがほとんどなく、コーダの大きな構造的変更のないこの適応手順は、符号化音声の主観的品質に著しい改善をもたらすことができる。出願人はまた、8キロビット/秒と16キロビット/秒との間の可変ビット速度で(低遅延)LD-CELPコーダに適用された図3の処理で良好な結果を得た。傾斜クラスは前述の場合と同一で、 $\Gamma_0=0.98$ ； $\lambda_0=4$ ； $\mu_0=1$ ； $\Delta_{\min,0}=0.6$ ； $\Delta_{\max,0}=0.8$ ； $\Gamma_1=0.98$ ； $\lambda_1=6$ ； $\mu_1=1$ ； $\Delta_{\min,1}=0.2$ ； $\Delta_{\max,1}=0.7$ であった。

【図面の簡単な説明】

【図1】本発明を実施することができるCELPデコーダの概略配置図である。

【図2】本発明を実施することができるCELPコーダの概略配置図である。

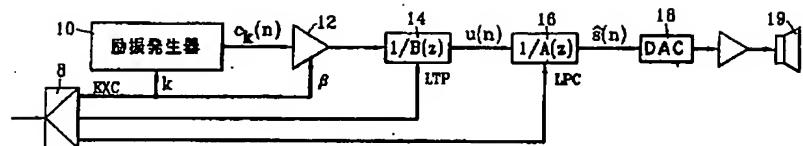
【図3】知覚重み付けを評価するための手順のフローチャート図である。

【図4】関数 $\log [(1-r)/(1+r)]$ のグラフを示す。

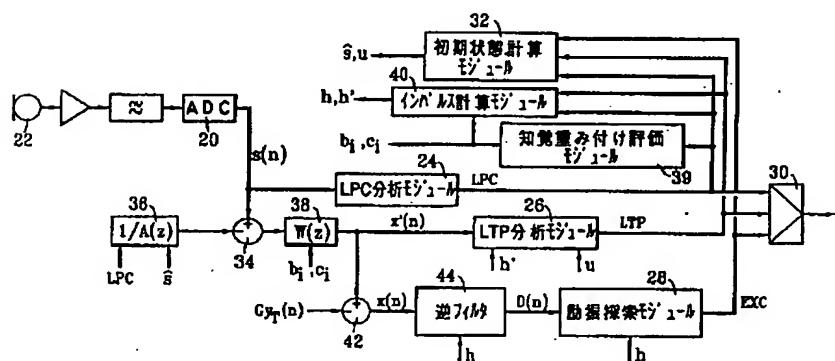
【符号の説明】

- 10 励振発生器
- 12 増幅器
- 14 長期合成フィルタ
- 16 短期合成フィルタ
- 20 アナログ/デジタル変換器
- 22 マイクロホン
- 24 分析モジュール
- 26 分析モジュール
- 28 分析モジュール

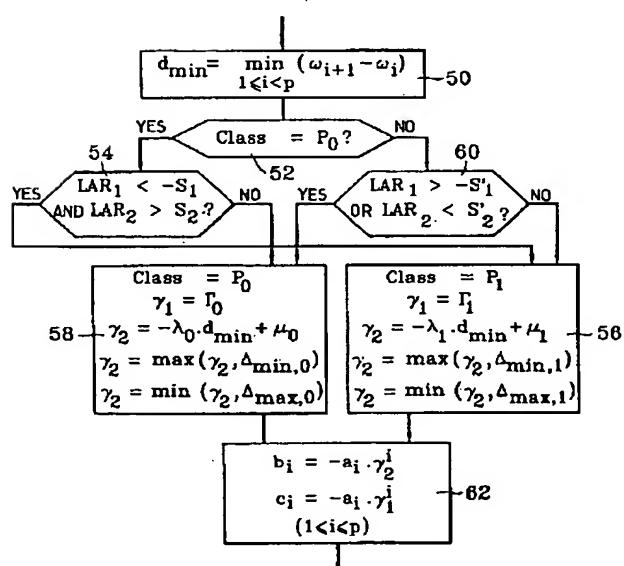
【図1】



【図2】



【図3】



【図4】

